Copy No. 10

DEFENCE R&D DÉFENSE

TRAINABLE LINEAR FILTERS FOR THE AUTOMATED DETECTION OF MINES IN SIDE SCAN SONAR IMAGES

R. T. Kessel

DEFENCE RESEARCH ESTABLISHMENT ATLANTIC

Technical Memorandum

DREA TM 1999-171

December 1999

National Defence Défense nationale

Canada

National Defence
Research and Development Branch
Défense nationale
Bureau de recherche et développement

TRAINABLE LINEAR FILTERS FOR THE AUTOMATED DETECTION OF MINES IN SIDE SCAN SONAR IMAGES

Ronald T. Kessel

December 1999

Approved by Ron Kuwahara:
Head/ Electro-Magnetics Section

TECHNICAL MEMORANDUM DREA 1999-171

Defence Research Establishment Atlantic
Centre de Recherches pour la Défense Atlantique

Canada

102-02564/

Abstract

The multibeam side scan sonar images used for sea mine hunting look rather like aerial photographs of the seafloor, but there are important differences. The sonar's automatic gain control, the consistent highlight-shadow signature progression of objects, and the absence of vanishing points, considerably reduce the degrees of freedom in an object's appearance, making sonar images better suited for linear detection schemes than optical photographs generally are. This paper demonstrates a trainable linear filter for the detection of mine-like objects in side scan sonar images. The training is fast and flexible, using actual or synthetic sonar images to prescribe what image patterns are to be detected and rejected, and with a sliding filter window of any shape to focus more particularly on important features. Measures of performance such as the probabilities of detection and false alarm, and the limits of trained discrimination and window size are derived. The method is demonstrated for mine-like targets in actual sonar images.

Les images de sonar à balayage latéral multi-faisceaux utilisées pour la chasse aux mines marines ressemblent plutôt à des photographies aériennes du plancher océanique, mais avec des différences importantes. La commande automatique de gain, la progression uniforme de signatures à surbrillance-masque des objets et l'absence de points disparaissant du sonar réduisent considérablement les degrés de liberté dans l'aspect d'un objet, ce qui rend les images sonar plus appropriées aux techniques de détection linéaire que les photographies optiques. Ce document fait la démonstration d'un filtre linéaire à capacité d'apprentissage pour la détection des objets ressemblant à des mines dans les images de sonar à balayage latéral. L'apprentissage est rapide et souple; il utilise des images sonar réelles ou synthétiques pour prescrire quelles configurations d'image doivent être détectées et rejetées, et se sert d'une fenêtre de filtrage mobile de n'importe quelle configuration, pour se concentrer plus particulièrement sur les caractéristiques importantes. Les mesures de performance, comme les probabilités de détection et de fausse alarme, et les limites de discrimination d'apprentissage et de taille de fenêtre sont obtenues par dérivation. La méthode est démontrée pour des cibles senblables à des mines dans des images sonar réelles.

DREA TM/99/171

Trainable linear filters for the automated detection of mines in side scan sonar images

Ronald T. Kessel

EXECUTIVE SUMMARY

Introduction

When side scan sonars are used in mine hunting, the sonar operator must recognize mines against a background of highly variable clutter that depends on the quality of the sonar, its altitude above the seafloor, and the nature of the seafloor. The images look like aerial photographs of the seafloor, but there are important differences. The sonar's automatic gain control, the consistent highlight-shadow signature progression of objects, and the absence of vanishing points, all reduce the degrees of freedom in an object's appearance, making sonar images more amenable to linear automated detection schemes than optical photographs generally are. This paper describes a linear filter for the automated detection of mines in sonar images, where a single filter is trained by a set of "snap-shots" representing the type of objects to be detected (mines) and rejected (mine-like rocks perhaps). The method is demonstrated for mine-like targets in actual sonar images.

Principal Results

The filter is constructed directly from the training images using Singular Value Decomposition (SVD). The training set must be self consistent in ways that are well defined. Analytic expressions for the probabilities of detection and false alarm are derived. The emphasis throughout is on practice, including the effects of signal noise and image mismatch, convenient performance curves, reasonable simplifying assumptions for mine hunting, and the limits of feasibility (the limits of trained discrimination, for instance, and the smallest training image size).

Significance of the Results

Analytic measures of performance (probabilities of detection and false alarm) make it possible to design for given specifications, or to establish when those specifications are too stringent for the linear filter. Performance trade-offs can be explored quantitatively by analysis for a given training set, without resorting to extensive trials or simulations of statistical performance.

Future Plans

The detection scheme can be focussed more particularly on features of given length scale and frequency content by applying the trainable filter in both the wavelet and Fourier transform domain of a sonar image, which may prove especially useful for texture classification and mapping of the seafloor. Several trainable filters might also be used to discriminate between different classes of objects, or between the features of a single object, rather like a neural network, now with a deterministic training method. The quantitative interplay between the trade-offs of the trainable filter might elsewhere serve to model their qualitative counterparts in human performance studies.

CONTENTS

1 Introduction **2**

2 The trainable linear filter **4**
2.1 Consistent detection and rejection training 5
2.2 Shift invariance 7
2.3 Design by Singular Value Decomposition (SVD) 8
2.4 The matched filter 9
2.5 Discriminating between similar images 9

3 Performance **11**
3.1 Probability of detection 11
3.2 Probability of *trained* false alarm 13
3.3 Probability of *untrained* false alarm 14
3.3.1 Augmented filter bank 16
3.4 Best detection-rejection discrimination 16
3.5 Minimum window size 17
3.6 Noise and allowable mismatch 18

4 Examples **20**
4.1 Mine target range 20
4.1.1 Example 1: Straightforward filter design 20
4.1.2 Example 2: Multiple training images 22
4.1.3 Example 3: Rejection training with non-rectangular windows 22
4.2 Example 4: Mine-like rocks 24

5 Conclusions and future work **28**

1. INTRODUCTION

When side scan sonars are used in mine hunting, the sonar operator must recognize mines against a background of highly variable clutter that depends on the quality of the sonar, its altitude above the seafloor, and on the nature of the seafloor—whether it is sand, rock, silt, sloping and so forth. Computer image processing is often used to help the operator identify mines, reject false targets, and to classify the seafloor by its visual texture. This paper describes an image processing scheme for object detection, and to a lesser extent bottom classification, that is based on linear filter design. The method is fast and flexible, using actual or synthetic sonar images to prescribe what image patterns are to be detected and rejected. More importantly, it is amenable to analysis to estimate its fundamental measures of performance—the probabilities of detection and false alarm—, making it possible to design for given performance specifications, or to establish when those specifications are too stringent for the linear approach.

A matched filter is perhaps the best known linear scheme [1][2]. Here the matched filter is generalized to the *trained* filter, where the filter is now designed to detect a number of patterns (different mines or viewing angles) and to *reject* other patterns (mine-like objects giving troublesome false alarms), with the matched filter resulting as a special case when just one training image is given for detection. The training images can be chosen freely, and may therefore entail inconsistencies that a linear filter cannot resolve. Ways of identifying and eliminating these contradictions, or managing near contradictions by way of performance trade-offs, are given here.

Non-linear processing methods such as neural networks have perhaps gained popularity over linear[3][4][5], but they are harder both to train and evaluate than linear schemes. Estimating the probabilities of detection and false alarm may in fact be the greatest burden for any new detection scheme, whether linear or non-linear, without which its performance remains questionable no matter how plausible it may at first appear. A thorough evaluation of the linear scheme is necessary in any case, if only to compare it against non-linear schemes to determine

which is better.

Side scan sonar images are monochromatic gray-scale images that lend themselves to linear processing more than photographic images generally do. The sonar's pulse and automatic gain control are designed to keep characteristic echo (highlight) from mine-like objects uniformly high across the swath, with highlights typically 10 dB higher than the noise level[6]. Some elementary detection schemes have been based on this energy principle alone [7][8]. The signature of an echo from objects proud of the seafloor, moreover, always progresses from highlight to receding shadow along a line perpendicular to the sonar tow path, Fig.(1). The shadow is variable in length depending on the altitude and range of the sonar, but it is never rotated or reversed with respect to the tow-path. And of equal importance, a sonar image, unlike conventional photography, is organized entirely on the basis of the travel time of the sonar pulse, from sonar to reflecting object, and back to the sonar, hence the distance to every object in a sonar image is known as a matter of course. The sonar image therefore has no vanishing perspective (no vanishing point). Rather, each pixel represents roughly the same area of the seafloor, typically 0.20 m by 0.20 m for mine hunting, so that physical dimensions of mine-like objects span roughly the same number of pixels no matter where they appear in the image (the shadows they cast of course become longer further from the sonar). All of these factors restrict the degrees of freedom in the patterns for detection in a sonar image considerably.

2. THE TRAINABLE LINEAR FILTER

Two-dimensional correlation, common in image processing, involves placing a rectangular window systematically over every pixel of the scanned image, and assigning a weighted sum of gray-scale intensities inside the window to a corresponding pixel in a new "filtered" version of the image. If the sliding window is R pixels high and C pixels wide, for instance, and the gray-scale intensity at each pixel of the scanned image falling inside the window is $M_{i,j}$, $i = 1 \ldots R$ and $j = 1 \ldots C$, then the *filter output* is

$$c = \sum_{i=1}^{R} \sum_{j=1}^{C} F_{i,j} M_{i,j}, \tag{2.1}$$

where $F_{i,j}$ are the filter coefficients. The operation is repeated with the window centered in turn over every pixel in the scanned image. The objective here, then, is to derive filter coefficients $F_{i,j}$ that will aid in the recognition (detection or rejection) of a number of $R \times C$ image patterns clipped from actual sonar images or produced synthetically using a computer model.

To this end, write the matrices M and F now as their one dimensional isomorphic counterparts, vectors $\mathbf{m}$ and $\mathbf{f}$, by placing the matrix rows end to end

$$\mathbf{m} = \left[\begin{array}{cccccccccc} M_{1,1} & M_{1,2} & M_{1,3} & \cdots & M_{1,C} & M_{2,1} & M_{2,} & M_{2,3} & \cdots & M_{R,C} \end{array} \right]^T, \tag{2.2}$$

$$\mathbf{f} = \left[\begin{array}{cccccccccc} F_{1,1} & F_{1,2} & F_{1,3} & \cdots & F_{1,C} & F_{2,1} & F_{2,2} & F_{2,3} & \cdots & F_{R,C} \end{array} \right]^T. \tag{2.3}$$

Then (2.1) becomes simply

$$c = \mathbf{m} \cdot \mathbf{f}. \tag{2.4}$$

Given a set of N *training* images M^n, $n = 1 \ldots N$, also $R \times C$ pixels in size and likewise rendered as vectors $\mathbf{m}_n$, and where some of these training images represent patterns to be detected and some rejected, then the simplest detection scheme is the binary,

$$c_n = \mathbf{m}_n \cdot \mathbf{f} = \begin{cases} 1 & \text{when } \mathbf{m}_n \in \mathcal{D} \text{ (detection)} \\ 0 & \text{when } \mathbf{m}_n \in \mathcal{R} \text{ (rejection).} \end{cases} \tag{2.5}$$

One could discriminate more particularly by assigning a different output c_n for different groups of training images, but the simple binary scheme will be followed here. The filter is *suboptimal* inasmuch as there is no effort here to maximize the likelihood function in Bayesian statistics as commonly done for a matched filter[2][1][9]. A *trained* filter should prove no less serviceable for not being optimal so long as it meets the given design specifications. It has the advantage of flexible trainability and straightforward analysis as we shall see.

Now collecting the pixel intensities of the N training images into the rows of a single matrix

$$\mathcal{M} = \begin{bmatrix} M^1_{1,1} & M^1_{1,2} & M^1_{1,3} & \cdots & M^1_{R,C} \\ M^2_{1,1} & M^2_{1,2} & M^2_{1,3} & \cdots & M^2_{R,C} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ M^N_{1,1} & M^N_{1,2} & M^N_{1,2} & \cdots & M^N_{R,C} \end{bmatrix} = \begin{bmatrix} \mathbf{m}_1^T \\ \mathbf{m}_2^T \\ \vdots \\ \mathbf{m}_N^T \end{bmatrix}, \tag{2.6}$$

and their respective prescribed filter outputs c_n into a vector

$$\mathbf{c} = \begin{bmatrix} c_1 & c_2 & \cdots & c_N \end{bmatrix}^T, \tag{2.7}$$

the detection scheme forms a linear system of equations

$$\mathcal{M}\mathbf{f} = \mathbf{c} \tag{2.8}$$

whose solution $\mathbf{f}$ holds the unknown trained filter coefficients.

The solution is complicated by the fact that $\mathcal{M}$ will generally be non-square and singular. If the number of training images N (rows in $\mathcal{M}$) is less than the number of pixels RC (columns in $\mathcal{M}$), $N < RC$, then the solution $\mathbf{f}$—if it exists at all—will be *under determined* and therefore non-unique. Or given the reverse, $N > RC$, the solution will be *over determined*, imposing excessive and potentially inconsistent constraints, and a solution $\mathbf{f}$ will almost certainly not exist. Choosing $N = RC$ would be an artificial limit on the training set that as a rule serves no better as we will see.

2.1. Consistent detection and rejection training

To illustrate an obviously ill posed filter design for which a solution filter design $\mathbf{f}$ does not exist, imagine specifying contrary objectives by including the same

training image twice: once $\mathbf{m}_1$ for detection, $c_1 = 1$, and once $\mathbf{m}_2 = \mathbf{m}_1$ for rejection, $c_2 = 0$. Clearly no solution exists because it is impossible for the filter to give two different outputs for the same input. Near contradictions of this kind may actually occur in mine hunting, where mines are to be detected but mine-like rocks are to be rejected; the two being similar in appearance. Less obvious contradictions may be struck unknowingly.

The *existence* of a solution $\mathbf{f}$ is a formal mathematical property established by fundamental theorems from the theory of linear operators[10] (if $\mathcal{M}$ is singular, then a solution $\mathbf{f}$ to (2.8) only exists if $\mathbf{c}$ is orthogonal to the null space of the adjoint operator $\mathcal{M}^T$), or again with equal rigor by theorems of linear algebra[11] (a solution $\mathbf{f}$ only exists if the rank of the augmented matrix $[\mathcal{M} \mid \mathbf{c}]$ is equal to the rank of $\mathcal{M}$), either of which can be turned to address the practical implications for linear filter design. A more direct approach builds on the elementary principle of linear dependence in the following way.

If the training images, rendered now as vectors $\mathbf{m}_k$, are linearly dependent—that is, one or more can be expressed as a linear combination of the others,

$$\mathbf{m}_k = \sum_{i \neq k} \alpha_i \mathbf{m}_i, \tag{2.9}$$

α_i being scalars—then it follows that the filter output c_k for a successful design must likewise be constrained

$$c_k = \mathbf{m}_k \cdot \mathbf{f} = \sum_{i \neq k} \alpha_i \mathbf{m}_i \cdot \mathbf{f} = \sum_{i \neq k} \alpha_i c_i. \tag{2.10}$$

If the prescribed values of c_k in the detection scheme (2.5) violate this constraint, then the filter is ill posed and has no solution.

Consider, for instance, a contradiction rather like that sighted above, when the same training image is included twice: first as $\mathbf{m}_1$ for detection, $c_1 = 1$, but now again in scaled form $\mathbf{m}_2 = \alpha \mathbf{m}_1$, also for detection, $c_2 = 1$. We are in this way asking the filter output to be invariant with respect to uniform scale changes in intensity, which may at times be a useful property. Notice, however, that the constraint (2.10) becomes $c_2 = \alpha c_1 = \alpha$ in contradiction with its prescribed value $c_2 = 1$ whenever $\alpha \neq 1$. Hence a linear filter always discriminates between a pattern and a level scaled version of that pattern in direct proportion to the scaling. It should be noted that the effects of scaling could be removed before linear processing by applying non-linear gray-scale level normalization, but this is not generally recommended for mine hunting.

Ill posed training therefore includes linearly dependent training images $\mathbf{m}_k$ with prescribed c_k that violate the constraint (2.10). The likelihood of inconsistency increases as the number of training images N increases. Indeed, if N is greater than the number of pixels RC in the training images, then there are no less than $N - RC$ linearly dependent training images (the number of rows in $\mathcal{M}$ exceeding the dimension of its domain by that amount), and it is virtually certain that the filter is ill posed. A general rule of thumb is that the number of training images N should be much less than the number of pixels in a training image, but this is no guarantee against inconsistency as we have seen for $N = 2$.

2.2. Shift invariance

It was shown that a linear filter always discriminates between a pattern and a level scaled version of that pattern. Yet it is possible to make a linear filter insensitive to uniform shifts (additive increase or decrease) in gray scale level in the following way. Let the uniform shift in scale be the additive scalar q. Then the filter output for an image $\mathbf{m}$ and its shifted version are

$$\begin{aligned} c &= \mathbf{m} \cdot \mathbf{f} \\ c' &= (\mathbf{m}+q\mathbf{e}) \cdot \mathbf{f}, \end{aligned} \tag{2.11}$$

respectively, where $\mathbf{e}$ has RC elements,

$$\mathbf{e} = [1, 1, \ldots 1]. \tag{2.12}$$

Clearly $c' = c$ for all q if $\mathbf{f}$ is perpendicular to $\mathbf{e}$, which can be enforced by including the vector $\mathbf{e}$ as a training image for rejection; i.e., $\mathbf{e} \cdot \mathbf{f} = 0$. The elements of $\mathbf{f}$ have zero mean because

$$\langle f_i \rangle = \frac{1}{RC} \mathbf{e} \cdot \mathbf{f} = 0. \tag{2.13}$$

Shift invariance is important not only to detect level-shifted versions of a given pattern equally well, but to reduce the number of false alarms as well. Any scanned image $\mathbf{m}$ can be written, for instance, as its mean intensity $\mu = \langle m_i \rangle$ plus a zero-mean variation $\mathbf{d}$,

$$\mathbf{m} = \mu\mathbf{e} + \mathbf{d}. \tag{2.14}$$

Viewing $\mathbf{d}$ as a zero-mean random vector uncorrelated with the filter $\mathbf{f}$, the expected filter output is then

$$\langle c \rangle = \langle \mathbf{m} \cdot \mathbf{f} \rangle = \mu RC \langle f_i \rangle + \langle \mathbf{d} \cdot \mathbf{f} \rangle = \mu RC \langle f_i \rangle, \tag{2.15}$$

which, because it is uncorrelated, should preferably be treated as a rejection, $\langle c \rangle = 0$. Shift invariance (2.13) ensures that this is true.

2.3. Design by Singular Value Decomposition (SVD)

Training entails the solution of a non-square singular system of equations (2.8). SVD is a robust numerical method for the purpose. SVD decomposes a matrix into the product of three[12][13]

$$\mathcal{M} = U\Lambda V^T \tag{2.16}$$

where U and V are orthonormal matrices, such that

$$\begin{aligned} UU^T &= I \text{ and } |\det(U)| = 1, \\ VV^T &= I \text{ and } |\det(V)| = 1; \end{aligned} \tag{2.17}$$

and Λ is a diagonal matrix holding the real positive singular values

$$\Lambda = \text{diag}\left[\begin{array}{cccc} \gamma_1 & \gamma_2 & \cdots & \gamma_{RC} \end{array}\right] \tag{2.18}$$

which we may assume here to be ordered from smallest γ_1 to largest γ_{RC}. The inverse of a non-singular matrix is

$$\mathcal{M}^{-1} = V^{-T}\Lambda^{-1}U^{-1} = V\Lambda^{-1}U^T. \tag{2.19}$$

A matrix is singular whenever one of more of its singular values are zero inasmuch as

$$|\det(\mathcal{M})| = |\det(U)|\,|\det(\Lambda)|\,|\det(V)| = \gamma_1 \cdot \gamma_2 \cdot \ldots \gamma_{RC} = 0. \tag{2.20}$$

The columns $\mathbf{v}_i$ of V whose same numbered singular values γ_i are zero form an orthonormal basis spanning the null-space $\mathcal{M}$. That is,

$$\mathcal{M}\mathbf{v}_i = \mathbf{0} \text{ and } \mathbf{v}_i \cdot \mathbf{v}_j = \begin{cases} 0 & \text{for} \quad i \neq j \\ 1 & \text{for} \quad i = j \end{cases}. \tag{2.21}$$

In practice, the inversion of a matrix suffers numerical instability when the smallest singular value is "small" compared with the largest—the ratio γ_{RC}/γ_1 being in fact a measure of the *condition* of the matrix[12]. It is not hard to show that the solution $\mathbf{f}$ would then ordinarily be very large and lie almost entirely in the null space. The runaway null space component can be suppressed by simply

zeroing the largest diagonal elements of Λ^{-1} whose same numbered singular values are very small. Substituting such a modified bounded inverse $(\Lambda^{-1})'$ into (2.19) regularizes the solution

$$\mathbf{f} = V\left(\Lambda^{-1}\right)' U^T \mathbf{c}. \tag{2.22}$$

This regularization invariably determines some vector $\mathbf{f}$, though it may not actually solve the original problem owing to inconsistencies in the training set (Section(2.1)) The solution f should therefore be tested by substitution back into the original equation (2.8) to confirm that it gives the desired outputs. Failure means that the filter design problem was ill posed from the outset. The contradiction can be eliminated by changing the training set $\mathcal{M}$ and computing a new filter $\mathbf{f}$, or by changing the detection rule (2.5)—by replacing $\mathbf{c}$ with $V\left(\Lambda^{-1}\right)' U^T \mathbf{c}$ for instance, giving a non-binary detection scheme. Of less importance, the solution $\mathbf{f}$ is also certainly not unique inasmuch as any linear combination of null space vectors plus $\mathbf{f}$ is also a solution. Thus any number of identically trained filters can be generated simultaneously, which may at times be useful (Section (3.3.1)).

It is important to note that the regularized solution (2.22) minimizes $|\mathbf{f}|$ in as much as it is orthogonal to the undetermined null-space component. This in turn optimizes the probability of detection as we shall see (Section (3.1)).

2.4. The matched filter

Given just one training image $\mathbf{m}_T$ for detection, moreover, the trained filter $\mathbf{f}$ then has no component orthogonal to the zero-mean version $\mathbf{d}_T$ of that training image (of the form (2.14)), hence $\mathbf{f}$ in (2.22) is the zero-mean *matched filter*

$$\mathbf{f} = \mathbf{d}_T / |\mathbf{d}_T|^2 . \tag{2.23}$$

2.5. Discriminating between similar images

More generally, the magnitude of the filter coefficients depends to a large extent on the similarity between patterns for detection and rejection. Assume, for instance, that a filter has been trained to detect a mine $\mathbf{m}_M$, but reject a similar mine-like rock $\mathbf{m}_R$. That is, in terms of the zero-mean (shift invariant) components for the

mine and rock

$$\mathbf{d}_R = \mathbf{d}_M + \boldsymbol{\delta}, \tag{2.24}$$

where $\boldsymbol{\delta}$ is a zero-mean difference between images non-parallel to $\mathbf{d}_M$ for consistency; i.e., $\boldsymbol{\delta} \cdot \mathbf{d}_M \neq 0$. The filter output for rejection is then

$$\mathbf{d}_R \cdot \mathbf{f} = 0, \tag{2.25}$$

from which it follows that

$$\begin{aligned} (\mathbf{d}_M + \boldsymbol{\delta}) \cdot \mathbf{f} &= 0, \\ \mathbf{d}_M \cdot \mathbf{f} + \boldsymbol{\delta} \cdot \mathbf{f} &= 0, \\ 1 + \boldsymbol{\delta} \cdot \mathbf{f} &= 0, \\ \boldsymbol{\delta} \cdot \mathbf{f} &= -1; \end{aligned} \tag{2.26}$$

and hence

$$|\boldsymbol{\delta}|\,|\mathbf{f}| \geq 1 \tag{2.27}$$

whereby

$$|\mathbf{f}| \geq \frac{1}{|\boldsymbol{\delta}|} = \frac{1}{|\mathbf{d}_M - \mathbf{d}_R|}. \tag{2.28}$$

The greatest lower bound on a successful filter design is therefore inversely proportional to the smallest difference between detection and rejection training images. Relatively large filter coefficients are therefore required to discriminate between very similar images for detection and rejection, and this affects the performance of the detector.

3. PERFORMANCE

Signal noise, environmental conditions, and variable sonar position on a sonar image invariably make objects in a scanned sonar image different from their counterparts in the training set. Hence the filter output for detection will differ somewhat from 1, and a *detection band* $1 \pm \varepsilon$ must be used,

$$1 - \varepsilon < c < 1 + \varepsilon; \text{ with } \varepsilon > 0. \tag{3.1}$$

Increasing ε increases the probability of detection, but it also increases the probability of false alarm as we shall see.

3.1. Probability of detection

The probability of detection is an important performance measure for any detection scheme. Here the *conditional probability of detection* P_D is defined as the probability that, given a mismatched version $\mathbf{m}_T + \mathbf{n}$ of a detection training image $\mathbf{m}_T$, the filter output will fall within the detection band (3.1),

$$1 - \varepsilon < (\mathbf{m}_T + \mathbf{n}) \cdot \mathbf{f} < 1 + \varepsilon; \tag{3.2}$$

or equivalently, that

$$-\varepsilon < \mathbf{n} \cdot \mathbf{f} < \varepsilon. \tag{3.3}$$

because $\mathbf{m}_T \cdot \mathbf{f} = 1$ in the binary detection scheme (2.5). The sum of many independent random variables tends to a normally distribution by the central limit theorem[14], hence the weighted sum $\mathbf{n} \cdot \mathbf{f}$ of n_i will as a rule be normally distributed with mean

$$\langle \mathbf{n} \cdot \mathbf{f} \rangle = \left\langle \sum_{i=1}^{RC} n_i f_i \right\rangle = RC \, \langle \langle n_i f_i \rangle \rangle = RC \, \langle n_i \rangle \, \langle f_i \rangle = 0 \tag{3.4}$$

(n_i and f_i being by assumption uncorrelated, and $\langle f_i \rangle = 0$ (2.13)); and statistical variance

$$\begin{aligned} \sigma^2 &= \left\langle (\mathbf{n}\cdot\mathbf{f} - \langle \mathbf{n}\cdot\mathbf{f}\rangle)^2 \right\rangle &&= \left\langle (\mathbf{n}\cdot\mathbf{f})^2 \right\rangle \\ &= \left\langle \sum_{i=1}^{RC}\sum_{j=1}^{RC} n_i n_j f_i f_j \right\rangle &&= \left\langle \sum_{i=1}^{RC} n_i^2 f_i^2 \right\rangle \\ &= RC\,\langle n_i^2\rangle\,\langle f_i^2\rangle &&= |\mathbf{n}|^2\,|\mathbf{f}|^2/(RC) \\ &= \sigma_N^2\,|\mathbf{f}|^2, \end{aligned} \tag{3.5}$$

where

$$\sigma_N^2 = |\mathbf{n}|^2/RC \tag{3.6}$$

is the variance of the noise and allowable mismatch. The probability of detection is therefore

$$P_D(\varepsilon) = \int_{-\varepsilon}^{\varepsilon} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} dx = \operatorname{erf}\left(\frac{\varepsilon}{\sqrt{2}\sigma}\right) = \operatorname{erf}\left(\frac{\varepsilon}{\sqrt{2}\sigma_N\,|\mathbf{f}|}\right) \tag{3.7}$$

where erf denotes the error function [15].

As expected, the probability of detection depends on 1) the quality of the training and scanned image through the mean noise and allowable mismatch variance σ_N^2, 2) the width of the detection band ε, and 3) the magnitude of the linear filter $|\mathbf{f}|$, which is minimized as a matter of course using SVD (Section (2.3)), thereby maximizing the probability of detection for a given training set.

The detection band width ε needed for a given filter $\mathbf{f}$, for a given probability of detection P_D, and for a given degree of noise and allowable mismatch σ_N^2, can be calculated from (3.7) using a root finder, as done to produce the performance curves in Fig.(2). The horizontal axis is expressed for convenience in terms of the *filter gain*

$$G(\mathbf{f},\mathbf{n}) = \frac{1}{\sigma^2} = \frac{RC}{|\mathbf{n}|^2\,|\mathbf{f}|^2}; \tag{3.8}$$

or equivalently

$$G(\mathbf{f},\sigma_N) = \frac{1}{\sigma_N^2\,|\mathbf{f}|^2}, \tag{3.9}$$

for which (3.7) becomes

$$P_D(\varepsilon) = \operatorname{erf}\left(\varepsilon\sqrt{\frac{G(\mathbf{f},\sigma_N)}{2}}\right). \tag{3.10}$$

More simply still, because the desired probability of detection is usually very high, $P_D \approx 0.995 \approx \text{erf}\,(2)$ for instance, it follows that

$$\varepsilon \approx 2\sqrt{\frac{2}{G\,(\mathbf{f},\sigma_N)}} = 2\sqrt{2}\sigma_N\,|\mathbf{f}|\,. \tag{3.11}$$

Increased effective noise σ_N therefore requires a proportionally increased detection band, and increased filter magnitude $|\mathbf{f}|$ results in proportionally increased sensitivity to noise.

For filters designed for both detection and rejection, the lower bound on $|\mathbf{f}|$ (2.28) implies that

$$\varepsilon > 2\sqrt{2}\frac{\sigma_N}{|\boldsymbol{\delta}|}, \tag{3.12}$$

where $|\boldsymbol{\delta}|$ is the smallest trained discrimination (2.24). Hence an ever wider detection band ε is required to compensate for ever smaller trained discrimination $|\boldsymbol{\delta}|$. On the other hand, given a single training image $\mathbf{m}_T$ of the form (2.14), the resulting matched filter (2.23) has magnitude $|\mathbf{f}| = |\mathbf{d}_T|^{-1}$, hence

$$\varepsilon \approx 2\sqrt{2}\frac{\sigma_N}{|\mathbf{d}_T|} = 2\sqrt{\frac{2}{RC}}\frac{\sigma_N}{\sigma_T}, \tag{3.13}$$

where $\sigma_T^2 = |\mathbf{d}_T|^2/RC$ is the variance of the lone training image, and the ratio $(\sigma_T/\sigma_N)^2$ is an effective signal-to-noise ratio and RC a gain.

3.2. Probability of *trained* false alarm

The *conditional probability of trained false alarm* P_{TF} is the probability that, given a slightly mismatched version a training image for rejection, the filter output nevertheless falls within the detection band,

$$1 - \varepsilon < \mathbf{n}\cdot\mathbf{f} < 1 + \varepsilon. \tag{3.14}$$

Proceeding much as before, we now find

$$P_{TF}\,(\varepsilon) = \int_{1-\varepsilon}^{1+\varepsilon} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} dx, \tag{3.15}$$

where σ^2 is the variance (3.5). This can be evaluated analytically,

$$P_{TF}(\varepsilon) = \frac{1}{2}\left[\mathrm{erf}\left(\frac{1+\varepsilon}{\sqrt{2}\sigma}\right) - \mathrm{erf}\left(\frac{1-\varepsilon}{\sqrt{2}\sigma}\right)\right], \tag{3.16}$$

which, to first order in $\varepsilon \ll 1$, is

$$P_{TF}(\varepsilon) \approx \varepsilon\sqrt{\frac{2}{\pi}}\frac{1}{\sigma}e^{-\frac{1}{2\sigma^2}} = \varepsilon\sqrt{\frac{2G(\mathbf{f},\sigma_N)}{\pi}}e^{-\frac{G(\mathbf{f},\sigma_N)}{2}} \tag{3.17}$$

The exact form (3.16) was used to produce the false alarm curves shown in Fig.(2). The asymptotic form (3.17) clearly shows that widening the detection band ε increases the probability of false alarm as expected. In view of (3.11) for a high probability of detection $P_D \approx 0.995$, the P_{TF} is in practice approximately

$$P_{TF}(\varepsilon) \approx \frac{4}{\sqrt{\pi}}e^{-\frac{G(\mathbf{f},\sigma_N)}{2}} = \frac{4}{\sqrt{\pi}}e^{-\frac{1}{2|\mathbf{f}|^2\sigma_N^2}}. \tag{3.18}$$

The performance determining factors are therefore the combined noise and allowed mismatch σ_N^2, and the magnitude of the filter $|\mathbf{f}|$. Increasing either increases the probability of trained false alarm. Recall that SVD minimizes $|\mathbf{f}|$ as a matter of course, thereby minimizing the probability of false alarm for a given training set.

3.3. Probability of *untrained* false alarm

False alarms also occur for images not among the training set, for which the filter output happens to fall within the detection band. To estimate their probability, assume first of all that the scanned image $\mathbf{m}$ does not resemble the training images. So far as the training images are concerned, then, the scanned image is entirely noise like, hence the probability density approaches a zero-mean normal distribution, with variance σ^2 given by (3.5) with the noise variance σ_N^2 replaced with the scanned image variance

$$\sigma_I^2 = \left\langle (m_{I_i} - \mu_I)^2 \right\rangle = \frac{|\mathbf{d}_I|}{RC}; \tag{3.19}$$

the scanned image being $\mathbf{m}_I = \mu_I\mathbf{e} + \mathbf{d}_I$ as in (2.14). The *conditional probability of untrained false alarm* is the probability that, given an image that is not a noisy

or mismatched version of one of the training images, the filter output falls within the detection band. That is,

$$P_{UF}(\varepsilon,\sigma') = \int_{1-\varepsilon}^{1+\varepsilon} \frac{1}{\sqrt{2\pi}\sigma'} e^{-\frac{x^2}{2\sigma'^2}} dx \tag{3.20}$$

where

$$\sigma' = \frac{|\mathbf{d}_I|\,|\mathbf{f}|}{\sqrt{RC}} = \sigma_I\,|\mathbf{f}|\,. \tag{3.21}$$

Figure (2) serves here again, showing P_{UF} as a function $G(\mathbf{f},\sigma_I)$ expressed in dB. The simplified estimate for trained false alarms (3.18) applies here again with the noise σ_N^2 replaced by the image variance σ_I^2,

$$P_{UF}(\varepsilon) \approx \frac{4}{\sqrt{\pi}} e^{-\frac{G(\mathbf{f},\sigma_I)}{2}} \approx \frac{4}{\sqrt{\pi}} e^{-\frac{1}{2|\mathbf{f}|^2\sigma_I^2}}. \tag{3.22}$$

Now the determining factors are the variance σ_I^2 of the scanned image and again the magnitude of the filter $|\mathbf{f}|$. Increasing either increases the probability of untrained false alarm.

To illustrate a realistic P_{UF} required for mine hunting, consider the following scenario[6]:

1. swath survey rate = 4 km^2 per hour;
2. pixel coverage = 0.2 m $\times$0.2 m per pixel;
3. allowable false alarm rate = 10 false alarms per hour;
4. typical mine size = 1.0 m $\times$ 0.5 m material cross sectional area, or 12.5 pixels per mine;

for which the maximum P_{UF} is

$$P_{UF} < 10 \text{ F.A. per hour} \times \frac{(0.2 \times 0.2)}{(4 \times 1000^2)} \text{ hours per pixel} = 10^{-7}. \tag{3.23}$$

3.3.1. Augmented filter bank

Untrained false alarms are easily identified as such by visual inspection because the scanned image triggering a false alarm bears little visual resemblance to any of the training images for detection. This suggests that several filters, much like neurons in a neural network, might be used to reduce false alarms by requiring that a detection be registered on all (or most) filters simultaneously before formally registering a detection. Such filters $\mathbf{g}_l$ could be constructed of the original filter $\mathbf{f}$ designed as above (2.22), plus a linear combination of null-field vectors $\mathbf{v}$ (2.21) returned by SVD,

$$\mathbf{g}_l = \mathbf{f} + \sum_i \beta_{li} \mathbf{v}_i. \tag{3.24}$$

It can be shown that all such filters satisfy the same training and detection scheme, though each would respond to particular instances of noise differently. Thus an untrained false alarm registered by one filter is unlikely to be registered by all. Preliminary analysis (not shown here) confirms that the probability of *untrained* false alarm can be dramatically reduced this way, but the probability of detection also apparently decreases in the proportion as the probability of *trained* false alarm decreases. The benefits for both detection and rejection is therefore questionable when rejection training is important. A filter bank may be warranted for detection training alone if P_{UF} is unavoidably high for a single filter.

3.4. Best detection-rejection discrimination

The lower bound on $|\mathbf{f}|$ (2.28) imposes a limit on the smallest feasible detection-rejection discrimination $|\boldsymbol{\delta}|$. Figure (3) plots that limit for typical probabilities of detection (3.10) and false alarm (3.20), as a function of the *variance-to-noise* ratio

$$VNR = 10 \log \left(\frac{|\mathbf{d}|^2}{|\mathbf{n}|^2} \right) = 10 \log \left(\frac{\sigma_I^2}{\sigma_N^2} \right). \tag{3.25}$$

The maximum possible VNR is roughly the signal-to-noise ratio (SNR) of the sonar, typically 12 dB perhaps, but it is usually smaller in practice because the effective noise variance σ_N^2 increases with allowances for mismatch (Section (3.6)). The smallest feasible *discrimination-to-variance* ratio along the vertical axis is

$$DVR = 10 \log \left(\frac{|\boldsymbol{\delta}|^2}{|\mathbf{d}|^2} \right) = 10 \log \left(\frac{\sigma_\delta^2}{\sigma_I^2} \right) \tag{3.26}$$

where $\sigma_\delta^2 = |\delta|^2 / RC$ is the smallest mean-square difference between detection and rejection training images. Note that the vertical axis includes a shift of $10 \log (RC)$ dB owing to the gain afforded by the window size RC, much as does in the filter gain (3.8). If $RC = 64$ pixels, for instance, then its contribution $10 \log (64) \approx 18$ dB must be subtracted from the level read from the vertical axis to get the corresponding minimum *discrimination-to-variance* ratio DVR. Assume, for example, that the filter is to perform with $P_{UF} \approx O(10^{-7})$, and VNR=12 dB. From Fig.(3), then, it would in principle be impossible to achieve detection-rejection discriminations less than about $15.5 - 18 = -2.5$ dB relative to the image variance (i.e., $\sigma_\delta^2 > 0.56\sigma_I^2$),without compromising the performance probabilities for the filter. More generally, the negative slope in Fig.(3) implies that *decreasing* VNR (due perhaps to decreasing sonar *signal-to-noise* ratio), *increases* the minimum discrimination level DVR. As expected, then, increased noise level makes it harder to discriminate between similar objects. Figure 3 shows the typical limits of detection-rejection discrimination for mine hunting applications.

Excessive probabilities of trained false alarm P_{TF} will also make rejection training ineffective. Figure (4) shows the P_{TF} as a function of variance-to-noise ratio VNR for typical probabilities of detection and *untrained* false alarm. P_{TF} becomes unacceptably large as the variance of (or information in) the scanned image becomes less than the noise and allowable mismatch, $VNR < 0$ dB.

3.5. Minimum window size

Discrimination generally improves as the training image size RC is increased, or more precisely, as the characteristic details of the objects to be discriminated are magnified or enhanced. Artificial expansion of the window, such as doubling its dimensions by simply replicating each pixel into a square of four pixels, is of course no use because it adds no additional information. Genuine training enhancement is required.

The smallest feasible window size can be estimated in the following way. The *trained* and *untrained* false alarms have the same approximate forms (3.18) and (3.22) (assuming high P_D), hence both may be treated simultaneously by dropping their distinguishing subscripts,

$$P_F \approx O\left(e^{-\frac{1}{2|\delta|^2\sigma^2}}\right) < O\left(10^{-n}\right); \tag{3.27}$$

the maximum prescribed probability in either case being fixed here by way of an exponent

$$n \leq \frac{\log(e)}{2|\mathbf{f}|^2 \sigma^2}. \tag{3.28}$$

Given a single training image for detection, $\mathbf{f}$ is the matched filter (2.23), and

$$n \leq \frac{\log(e)}{2} \frac{|\mathbf{d}_T^2|}{\sigma^2} \leq RC \frac{\log(e)}{2} \frac{\sigma_T^2}{\sigma^2}, \tag{3.29}$$

whereby

$$RC \geq \frac{2n}{\log(e)} \frac{\sigma^2}{\sigma_T^2} \approx 4.605n \frac{\sigma^2}{\sigma_T^2}. \tag{3.30}$$

n and σ^2 are taken for either false alarm, *trained* or *untrained*, which ever maximizes the lower bound for RC. Choosing the *untrained* false alarm, for instance, while assuming significant clutter in the scanned image such that the image variance $\sigma_I^2 \rightarrow \sigma^2$ roughly equals that of the training image σ_T^2, and assuming a typical probability of false alarm with $n = 7$ as in (3.23), we find that

$$RC > 4.605n \approx 32 \text{ pixels}. \tag{3.31}$$

In practice, then, we would expect to use window sizes somewhat larger than 32 pixels in mine hunting, though smaller could be used in benign environments with little clutter, where the background variance σ_I^2 falls considerably below than that of the training image σ_T^2 (see Example 3, Section (4.1.3) below).

3.6. Noise and allowable mismatch

The term σ_N^2 includes both the image noise (random perturbation away from the "true" sonar echo return owing to signal noise, etc.) and the allowable mismatch (difference between the training image of an object and its appearance in the scanned image due primarily to changing perspectives). In practice the noise can be estimated from deep shadows in the scanned sonar image, setting it equal to the variance σ_{shadow}^2 of the gray-scale intensity in the shadows. The noise distribution need not be Gaussian (see Section (3.1)).

The allowable mismatch is more problematic. One could imagine collecting sonar images $\mathbf{m}_s = \mu_s \mathbf{e} + \mathbf{d}_s$ of the same object from different viewing angles and ranges, and then computing an expected error (or noise) power for the whole set,

$$\sigma_{\text{mismatch}}^2 = \left\langle \frac{|\mathbf{d}_T - \mathbf{d}_s|^2}{RC} \right\rangle. \tag{3.32}$$

This is demonstrated below in Example 2, Section(4.1.2), using a set of two training images—reversed perspectives of the same mine. Alternatively, the set of perspectives $\mathbf{m}_i$ might be generated using computer simulation, perhaps on the basis of a simple highlight-shadow model of the object. There are many other ways of assessing the allowable mismatch for a given training set, but none will suit all possible training sets equally well.

If the performance of the automated pattern recognition is to duplicate that of a human sonar operator, then allowance should also be made for *visual* mismatch. Visual mismatch accounts for the fact that the linear processor responds proportionally to the smallest increments in gray-scale level, whereas the human eye distinguishes less precisely. Simple experiments with the visual perception of 256 (8-bit) gray-scale levels show that, when presented with the entire scale of levels changing uniformly from lowest 0 to highest 255, the human eye views roughly 10 consecutive gray-scale levels as one apparent level, giving is a perceptual uncertainty of about

$$\sigma_{\text{visual}} \approx 10. \tag{3.33}$$

The total noise and allowable mismatch is the incoherent sum of all three

$$\sigma_N^2 = \sigma_{\text{noise}}^2 + \sigma_{\text{mismatch}}^2 + \sigma_{\text{visual}}^2. \tag{3.34}$$

The total is bounded by the span of the image gray-scale

$$\sigma_N^2 < (2^n - 1)^2, \tag{3.35}$$

where n is here the number of bits representing the image intensity; typically $n = 8$. The upper bound might in fact be used on occasion, provided that it is not further constrained by the probabilities of false alarm. When rejection training is used, it follows from both the approximate P_{TF} (3.18) and the lower bound on $|\mathbf{f}|$ in (2.28) that

$$\sigma_N < \frac{1}{|\mathbf{f}|}\sqrt{\frac{-\log(e)}{2\log(P_{TF})}} < |\boldsymbol{\delta}|\sqrt{\frac{-\log(e)}{2\log(P_{TF})}} = \sigma_\delta\sqrt{\frac{-RC\log(e)}{2\log(P_{TF})}}, \tag{3.36}$$

in which P_{TF} assumes its prescribed maximum value. If $RC = 32$ pixels, for instance, and $P_{TF} = 10^{-7}$ for reliable rejection, then the quantity under the square root is close to unity. That is, the noise and allowable mismatch must as a rule be somewhat less than the trained discrimination, which is much as we would expect for human pattern recognition.

4. Examples

4.1. Mine target range

The Defence Research Establishment Pacific (DREP) set up a sonar test range in ocean waters near Esquimault, British Columbia, where a number of "mine-like" concrete pipes were placed on the ocean floor. Sonar images taken from the range provide a realistic demonstration of the linear filter for mine-hunting applications in a benign low-clutter environment. Figure (5) is a multi-beam side-scan sonar image the entire target range (8-bit, 0-255 gray-scale levels). For clarity, the set of complete targets—those whose highlights and shadows are entire—have been clipped and pasted in an enlarged composite Fig.(6).

4.1.1. Example 1: Straightforward filter design

Our first inclination may be to proceed as follows:

1. select one representative mine $M1$ for training (see Fig.(7)),
2. design a (shift invariant, matched) filter trained for detection of that mine as described in Section (2.3),
3. scan the image by taking the two-dimensional correlation with that filter,
4. set the detection band width ε to achieve a prescribed probability of detection, see Table (4.1),
5. produce a *detection overlay*, a replica of the scanned image with pixels "whited out" (set to maximum gay-scale level) where the correlation fell within the detection band to register a detection (see Fig.(7)),
6. verify that the probability of untrained false alarm is small, see Table (4.2).

Parameter	Value
window size (pixels)	$RC = 12 \times 22 = 264$
filter magnitude	$\|\mathbf{f}\| = 9.459 \times 10^{-4}$
allowed detection mismatch	$\sigma_{\text{mismatch}} = 0$
allowed visual mismatch	$\sigma_{\text{visual}} = 10$
noise (from shadow of $M1$)	$\sigma_{\text{noise}} = 18.10$
total noise and allowed mismatch	$\sigma_N = 21.13$
filter gain (dB)	$G(\mathbf{f},\sigma_N) = 34.17$
probability of detection (design specification)	$P_D = 0.990$
detection bandwidth	$\varepsilon = 0.050$

Table 4.1: Parameters relevant to computing the detection band width for Example 1.

Parameter	Value
scanned image variance	$\sigma_I = 35.92$
filter gain (dB)	$G(\mathbf{f},\sigma_I) = 29.38$
variance-to-noise ratio (dB)	$VNR = 5.95$
untrained probability of false alarm	$P_{UF} \ll 10^{-18}$

Table 4.2: Parameters relevant to the probability of untrained false alarm for Example 1.

The filter is a matched filter (2.23) because there was only one training image. One perfect detection $c = 1$ is registered in the overlay in Fig.(7), centered on the original training image $M1$. The detection is too perfect in fact, for no other targets were detected because the filter was trained to discriminate too particularly, by a large and detailed image with no allowance for mismatch beyond the noise and visual, $\sigma_{\text{mismatch}} = 0$.

Note that the probability of untrained false alarm P_{UF} is extremely small. Indeed, as the entire target range is about 2200 $\times$500 pixels large, the probability of encountering a false alarm in the target range is very much less than $2200 \times 500 \times 10^{-18} = O(10^{-12})$. No false alarms are in fact registered when the filter is applied to the entire range (not shown here). We may therefore increase the allowed mismatch considerably, and generalize the training without incurring excessive false alarms.

4.1.2. Example 2: Multiple training images

To generalize the training, let us

1. smooth the training image $M1$ (3×3 sliding average),
2. add a another image to the training set: a flipped version of this smoothed $M1$ about its horizontal axis, to simulate a change in the perspective without ruining the highlight-shadow progression, and
3. set the allowable mismatch to roughly accommodate both the range of all perspectives between $M1$ and its flipped version, and a 10 % variation in the strength of a highlight by setting

$$\begin{aligned} \sigma_{\text{mismatch}} &= (1 + 0.1) \times \text{standard deviation of } M1 \text{ minus its flipped version} \\ &= 1.1 \times 53.16 = 58.48 \end{aligned} \tag{4.1}$$

which obviously dominates other contributions to the noise and mismatch allowed above.

Key parameters for this example are summarized in Table (4.3). Figure (8) shows the detection overlay registering detections of all mine-like objects. The probability of false alarm P_{UF} remains more than adequately low. The false detections along the left edge are forgiven here because they are due to edge-effects in the two-dimensional correlation, where only part of the filter window has been used in the correlation.

4.1.3. Example 3: Rejection training with non-rectangular windows

The preceding examples concentrated on the large mine-like targets while ignoring the smaller (targets 9 and 13 in Fig.(6)). Here we demonstrate rejection training by detecting the smaller mines while rejecting larger, as well as *non-rectangular* filter windows that focus more particularly on the shadow and highlight regions alone.

Masking for non-rectangular windows The effect of the rectangular $R \times C$ pixel filter window considered thus far can be made equivalent to that of a smaller window having any shape using the principle of shift invariance (Section (2.2)) together with *masking*—the elimination of unwanted portions of a rectangular

Parameter	Value
window size (pixels)	$RC = 12 \times 22 = 264$
filter magnitude	$\|\mathbf{f}\| = 1.201 \times 10^{-4}$
allowed detection mismatch	$\sigma_{\text{mismatch}} = 58.48$
allowed visual mismatch	$\sigma_{\text{visual}} = 10$
noise (from shadow of $M1$)	$\sigma_{\text{noise}} = 18.10$
total noise and allowed mismatch	$\sigma_N = 62.03$
filter gain (dB)	$G(\mathbf{f},\sigma_N) = 22.56$
probability of detection (design specification)	$P_D = 0.990$
detection bandwidth	$\varepsilon = 0.192$
scanned image variance	$\sigma_I = 35.92$
filter gain (dB)	$G(\mathbf{f},\sigma_I) = 27.30$
untrained probability of false alarm	$P_{UF} \ll 10^{-18}$

Table 4.3: Parameters relevant to Example 2.

training image by setting their gray-scale level to the mean of all portions retained. Shift invariance ensures that the filter is insensitive to these mean-masked portions; the filter coefficients being zero (or very small) there. The filter gain $G(\mathbf{f},\sigma)$ in (3.9) is written in a form that accommodates the change in the effective window size for the probability estimates.

Figure (9) shows the masked version of $M2$ used for detection training in this example, where the original rectangular image (16×21 pixels) was thresholded to retain the highlight region (levels above 150; 20 pixels in all) and shadow (below 40; 13 pixels in all) regions, while the surrounding portions are effectively discarded, so far as the shift-invariant filter is concerned, by setting their level to the mean of the retained highlight-shadow regions (equal to 116.4; 303 pixels in all), This prescribed mean level equals the average level of the entire masked rectangular window.

When focussing on such relatively small distinguishing features this way, the filter is much less sensitive to changes in the orientation of the object, and to the details of the dead zone between the highlight and shadow and surrounding regions. Hence the allowable mismatch must be reduced accordingly . Here we allow for noise and visual mismatch only, setting $\sigma_{\text{mismatch}} = 0$.

Rejection training The linear filter is also insensitive to the along-track length of the object (vertical length in Fig.(9)), causing long track-parallel objects to

Parameter	Value
window size (pixels)	$RC = 16 \times 21 = 336$
non-rectangular training window size (pixels)	33
filter magnitude	$\|\mathbf{f}\| = 2.505 \times 10^{-4}$
allowed detection mismatch	$\sigma_{\text{mismatch}} = 0$
allowed visual mismatch	$\sigma_{\text{visual}} = 10$
noise (from shadow of $M1$)	$\sigma_{\text{noise}} = 18.10$
total noise and allowed mismatch	$\sigma_N = 20.69$
filter gain (dB)	$G(\mathbf{f},\sigma_N) = 25.71$
probability of detection (design specification)	$P_D = 0.990$
detection bandwidth	$\varepsilon = 0.133$
trained probability of false alarm	$P_{TF} \ll O(10^{-18})$
scanned image variance	$\sigma_I = 35.92$
filter gain (dB)	$G(\mathbf{f},\sigma_I) = 20.19$
untrained probability of false alarm	$P_{UF} \ll 10^{-18}$
variance-to-noise ratio (dB)	$VNR = 4.79$
smallest trained discrimination	$\|\boldsymbol{\delta}\| = 964.46$

Table 4.4: Parameters relevant to Example 3.

register many sequential detections side by side. Rejection training is used here to suppress the detection of the larger mines, with three rejection images in all: 1) a smoothed version of $M1$, 2) the same flipped about its horizontal axis, and 3) a smoothed version of $M3$; where all of these have been individually masked using the same highlight and shadow thresholds as for the training image. All training images, the final filter, and the detection overlay are shown in Fig.(9). Key parameters are in Table (4.4). The two smaller mines but none of the larger mines are detected as planned. It is found that without rejection training, or without any one of the three rejection training images, detections are registered along the partial length of several of the larger mine-like targets.

4.2. Example 4: Mine-like rocks

Figure (10) shows a more troublesome rocky seafloor, not part of the target range, with no mines present. We proceed, then, with two applications in mind: 1) mine hunting among so many mine-like rocks, where rejection training may be especially useful; and 2) estimating mine-like rock populations to classify the

seafloor according to its potential for hiding mines among clutter. Our approach is to train a filter to detect (or count) first small rocks, and then large rocks, which may alternately be considered as mine-like objects. Two training images (Fig.(10)) are used with masking to focus on highlight-shadow signatures, first the small rock for detection and the large for rejection, and then the large for detection and small for rejection. Figure (10) shows the relative magnitudes of the filter coefficients below their respective detection overlays. Key parameters for each are summarized in Table (4.5) and (4.6), respectively.

The first filter detected 25 small rocks, and the second 24 large (ignoring edge effects along the left edge of the overlay). Notice that $P_{UF} = 10^{-4.84}$ in Table (4.5) when detecting small rocks would ordinarily be unacceptably large; the problem being that the 16 operative pixels in the non-square detection training window is too small.. When multiplied by the number of pixels in the scanned image, it translates into a prediction of one (0.96) false alarm in the image.

We may or may not be satisfied with the filters' detection and discrimination, arguing justifiably, for instance, that some small rocks were over looked and that some large rock detections were in fact uplifted ridges and not individual rocks. Indeed, the designer's subjective inclinations for training and allowed mismatch shape the filter performance to a large degree. Different training images and allowable mismatch would result in slightly different inclinations in detection; inclinations that are not completely addressed by the usual probabilities of detection and false alarm. It is not surprising, however, that the discrimination of a trainable detection scheme should depend on its training this way. This may at times be an advantage for it allows an operator (or researcher) to tailor the filter to suit his or her expert judgement and requirements.

Parameter	Value
window size (pixels)	$RC = 13 \times 13 = 169$
non-rectangular training window size (pixels)	16
filter magnitude	$\|\mathbf{f}\| = 4.512 \times 10^{-4}$
allowed detection mismatch	$\sigma_{\text{mismatch}} = 0$
allowed visual mismatch	$\sigma_{\text{visual}} = 10$
noise (from shadow)	$\sigma_{\text{noise}} = 32.59$
total noise and allowed mismatch	$\sigma_N = 34.09$
filter gain (dB)	$G(\mathbf{f},\sigma_N) = 16.26$
probability of detection (design specification)	$P_D = 0.990$
detection bandwidth	$\varepsilon = 0.396$
trained probability of false alarm	$P_{TF} < 10^{-8}$
scanned image variance	$\sigma_I = 46.11$
filter gain (dB)	$G(\mathbf{f},\sigma_I) = 13.64$
untrained probability of false alarm	$P_{UF} = 10^{-4.84}$
variance-to-noise ratio (dB)	$VNR = 2.62$
smallest trained discrimination	$\|\delta\| = 367.41$

Table 4.5: Parameters relevant to Example 4, when trained to detect small rocks while rejecting large.

Parameter	Value
window size (pixels)	$RC = 13 \times 13 = 169$
non-rectangular training window size (pixels)	42
filter magnitude	$\|\mathbf{f}\| = 2.722 \times 10^{-3}$
allowed detection mismatch	$\sigma_{\text{mismatch}} = 0$
allowed visual mismatch	$\sigma_{\text{visual}} = 10$
noise (from shadow)	$\sigma_{\text{noise}} = 32.59$
total noise and allowed mismatch	$\sigma_N = 34.09$
filter gain (dB)	$G(\mathbf{f},\sigma_N) = 20.65$
probability of detection (design specification)	$P_D = 0.990$
detection bandwidth	$\varepsilon = 0.239$
trained probability of false alarm	$P_{TF} < 10^{-24}$
scanned image variance	$\sigma_I = 46.11$
filter gain (dB)	$G(\mathbf{f},\sigma_I) = 18.03$
untrained probability of false alarm	$P_{UF} < 10^{-12}$
variance-to-noise ratio (dB)	$VNR = 2.62$
smallest trained discrimination	$\|\delta\| = 367.41$

Table 4.6: Parameters relevant to Example 4, when trained to detect large rocks while rejecting small.

5. Conclusions and future work

A linear filter can be trained to detect some objects and reject others. The filter's performance can be evaluated directly through analytic expressions for the probabilities of detection and false alarm. These in turn cast the usual performance trade-offs quantitatively, including especially the trade-off between detection and false alarm (increasing the likelihood of one necessarily increasing that of the other), and between discrimination and error (the greater the likeness between objects for detection and rejection, the greater the likelihood of error). The filter training is flexible inasmuch as the images for training can be clipped directly from actual sonar images or generated synthetically using computer models, and they may be converted to equivalent non-rectangular training windows that focus more particularly on the features of an object. It was shown that the training set must be self consistent in ways that are well understood, and how a contradiction in a training set can be identified and corrected. The training is computationally efficient inasmuch as the filter coefficients are determined directly from the training images using Singular Value Decomposition (SVD), with any number of filters trained at once from a single training set, which together reduce the probability of false alarms.

Projecting somewhat further, trainable linear filters might also be used as one component in a non-linear pattern recognition scheme in which non-linear preprocessing is applied to the images before linear filtering. They could also be applied without change in the wavelet or Fourier transform domain of a sonar image, to focus the detection scheme on features of a particular length scale or frequency content. The Fourier transform domain may prove especially useful for texture classification and segmentation of the seafloor. All three domains (original image, wavelet, and Fourier) might be scanned simultaneously, each with its own specially trained filter, to improve detection and classification. Several trainable filters might also be used to discriminate between different classes of objects in a sonar image, with each filter trained by appropriate representatives from a given class. A single training image itself could be subdivided into a number of smaller

sub-windows, with each sub-window having its own linear filter to trained to detect particular features of a single object to classify them. Using many filters in these ways would be rather like using a neural network, but now with a straightforward deterministic training method. Finally, the quantitative interplay of the trainable filter's trade-offs might elsewhere serve as a model of their qualitative counterparts in human performance studies.

BIBLIOGRAPHY

[1] W.K. Pratt, *Digital Signal Processing*, 2nd Ed., John Wiley and Sons, New York, 1991.

[2] M.P. Ekstrom, Editor, *Digital Image Processing Techniques*, Academic Press, 1984.

[3] G.J. Dobeck, J.C. Hyland, and L. Smedly, "Automated detection/classification of sea mines in sonar imagery," Naval Research Reviews, Office of Naval Research, V.49, pp. 9-20, 1997.

[4] J.C. Hyland and G.J. Dobeck, "Sea mine detection and classification using side-looking sonar," *Proceedings of the SPIE Conference on Detection Technologies for Mines and Minelike Targets*, 17-21 April 1995, Orlando Florida, V. 2496, pp. 466-474.

[5] R.P. Gorman and T.J. Sejnowski, Learned classification of sonar targets using a massively parallel network, *IEEE Transactions on Acoustics Speech, and Signal Processing*, Vol. 36, 1988.

[6] J. Paul, P. Johnson, and J. Harris, "*Final report for the Canadian remote minehunting sonar study*," Canadian National Defence Research and Development Branch Contractor Report DREA CR/96/451A, 1996.

[7] R.J. Urick, *Principles of underwater sound for engineers*, McGraw Hill, New York, 1967.

[8] A.D. Waite, *Sonar for practising engineers*, 2nd Ed, Thomson Marconi Sonar Ltd, Great Britain, 1998.

[9] B.R. Calder, L.M. Linnett, and D.R. Carmichael, A Bayesian approach to object detection in side scan sonar, IEEE Proceedings of Vision, Image and Signal Proc. v. 145, 1998.

[10] R. Courant and D. Hilbert, *Methods of mathematical physics Vol. 1*, Interscience Publishers, New York, 1953.

[11] I.S. Sokolnikov and E.S. Sokolnikov, *Higher mathematics for engineers and scientists*, 2nd Ed., McGraw Hill, New York, 1941.

[12] G.H Golub and C. Van Loan, *Matrix Computations*, John Hopkin University Press, 1983.

[13] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vettering, *Numerical Recipes*, Cambridge University Press, Cambridge, 1986.

[14] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd edition, McGraw Hill, 1991.

[15] M. Abramowitz and I.A. Stegun, *Handbook of mathematical functions*, National Bureau of Standards, Applied Mathematics Series 55, Washington DC, 1965.

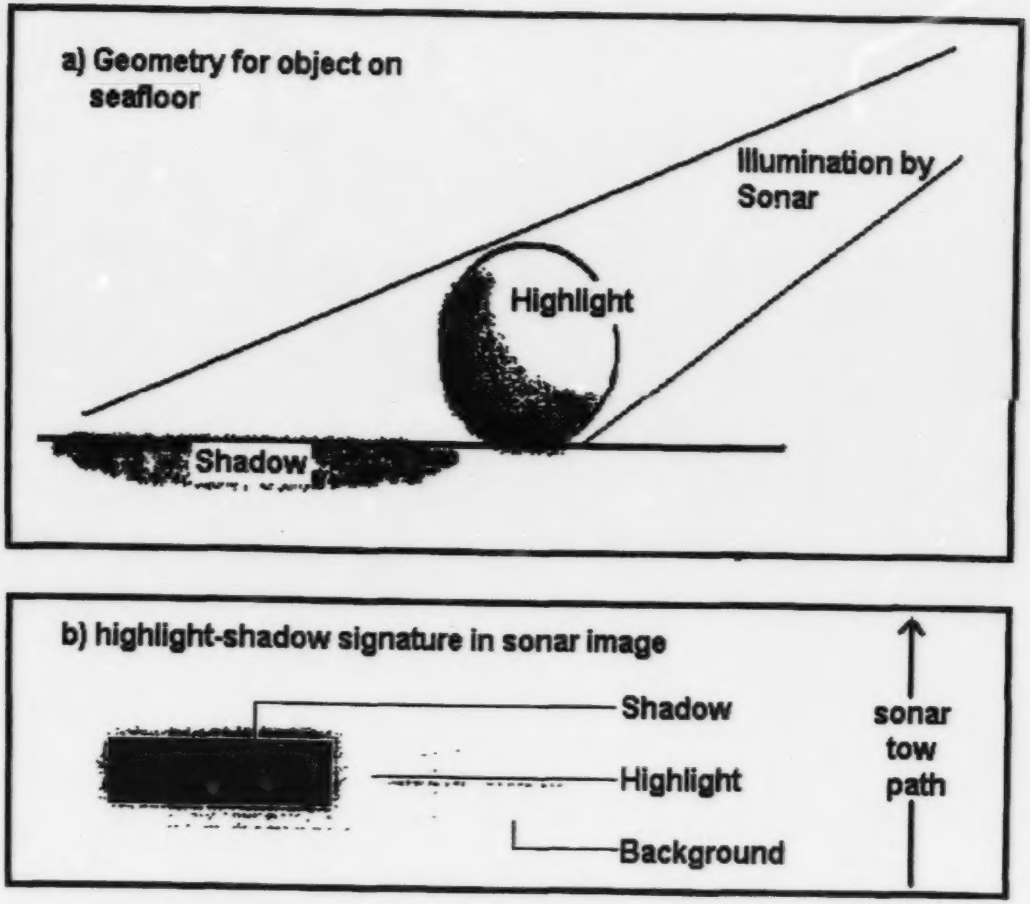

Fig.1 Typical highlight-shadow signature in side scan sonar image for an object proud of the seafloor.

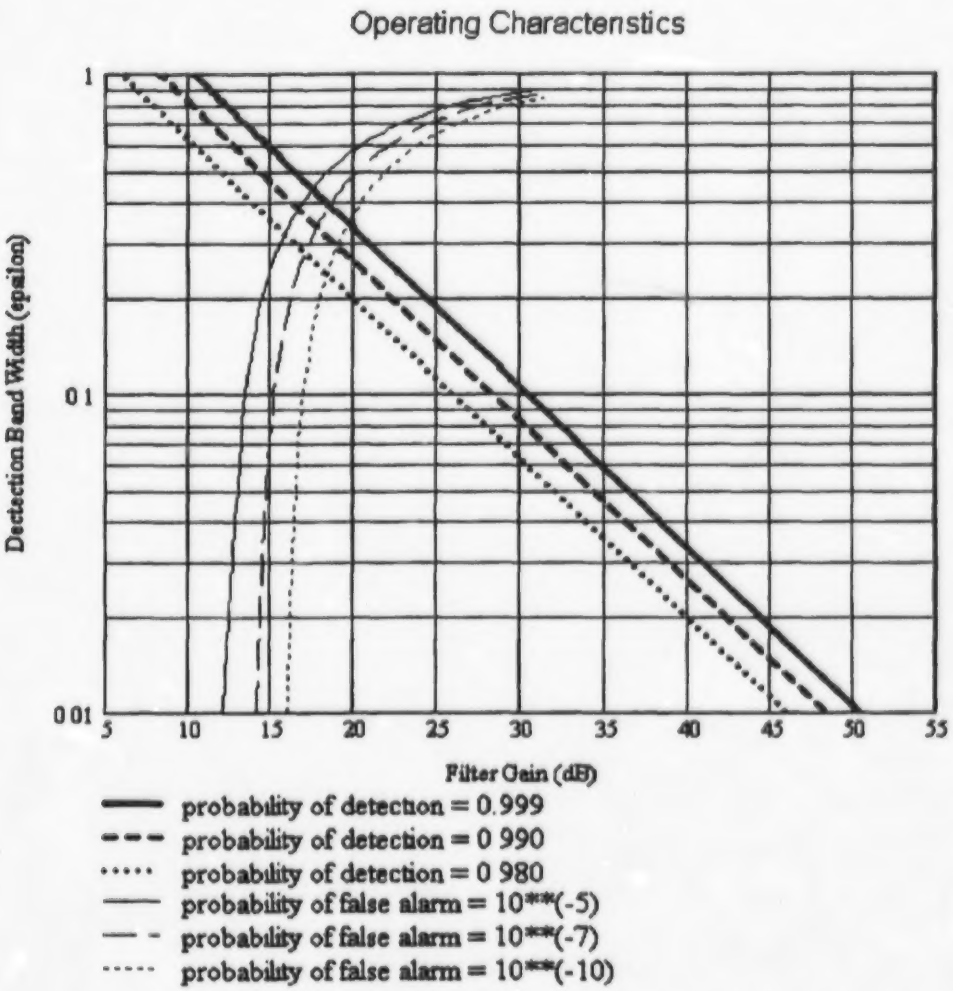

Fig.2 Operating characteristics for a linear filter. The allowed noise and mismatch σ_N determines the filter gain $G(f,\sigma_N)$ (37), which in turn determines the width of the detection band ε for a given probability of detection P_D (heavy lines). This ε, together with the same filter gain $G(f,\sigma_N)$, determines the probability of *trained* false alarm P_{TF} (thin lines); or together with the alternate filter gain $G(f,\sigma_I)$, determines the probability of *untrained* false alarm P_{UF} (thin lines again), σ_I^2 being the variance of the scanned image.

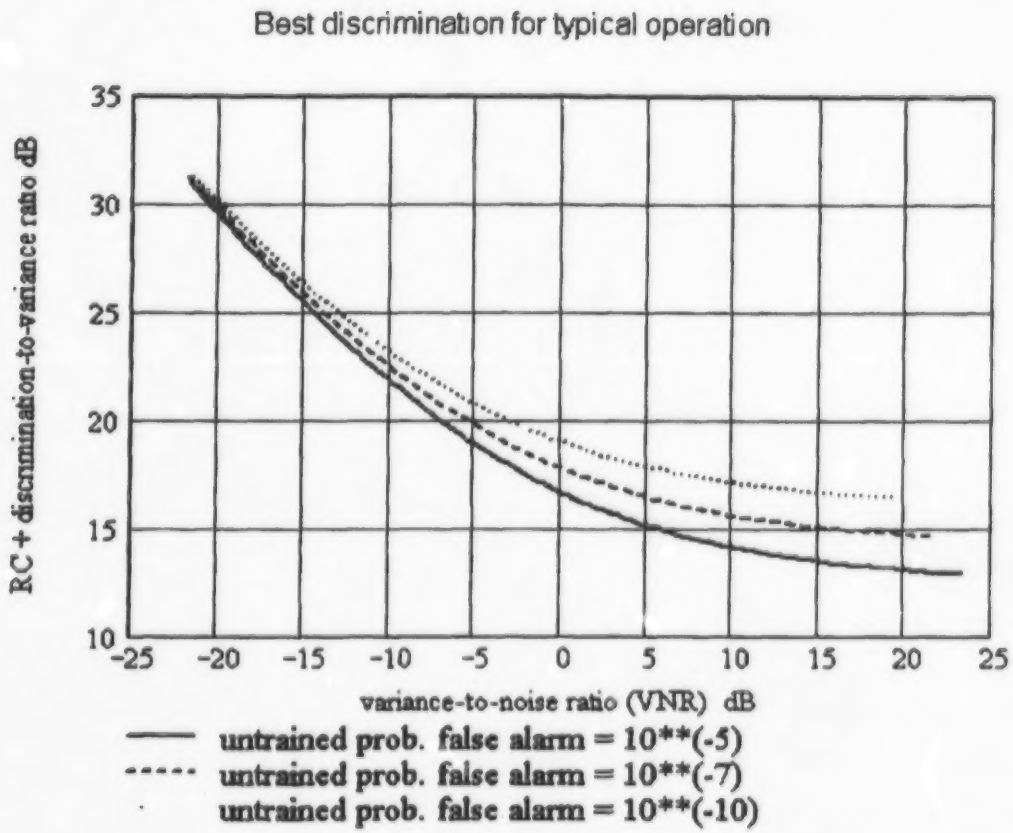

Fig. 3 Smallest detection-rejection discrimination
The smallest possible trained discrimination, expressed here as the discrimination-to-noise ratio (*DNR*) in (55), increases as the scanned image variance-to-noise ratio (*VNR*) in (54) decreases. The curves were computed assuming a probability of detection P_D=0.990, and three typical probabilities of *untrained* false alarm as noted in the legend. *VNR* is roughly proportional to the sonar signal-to-noise ratio (*SNR*) when there is no allowed mismatch.

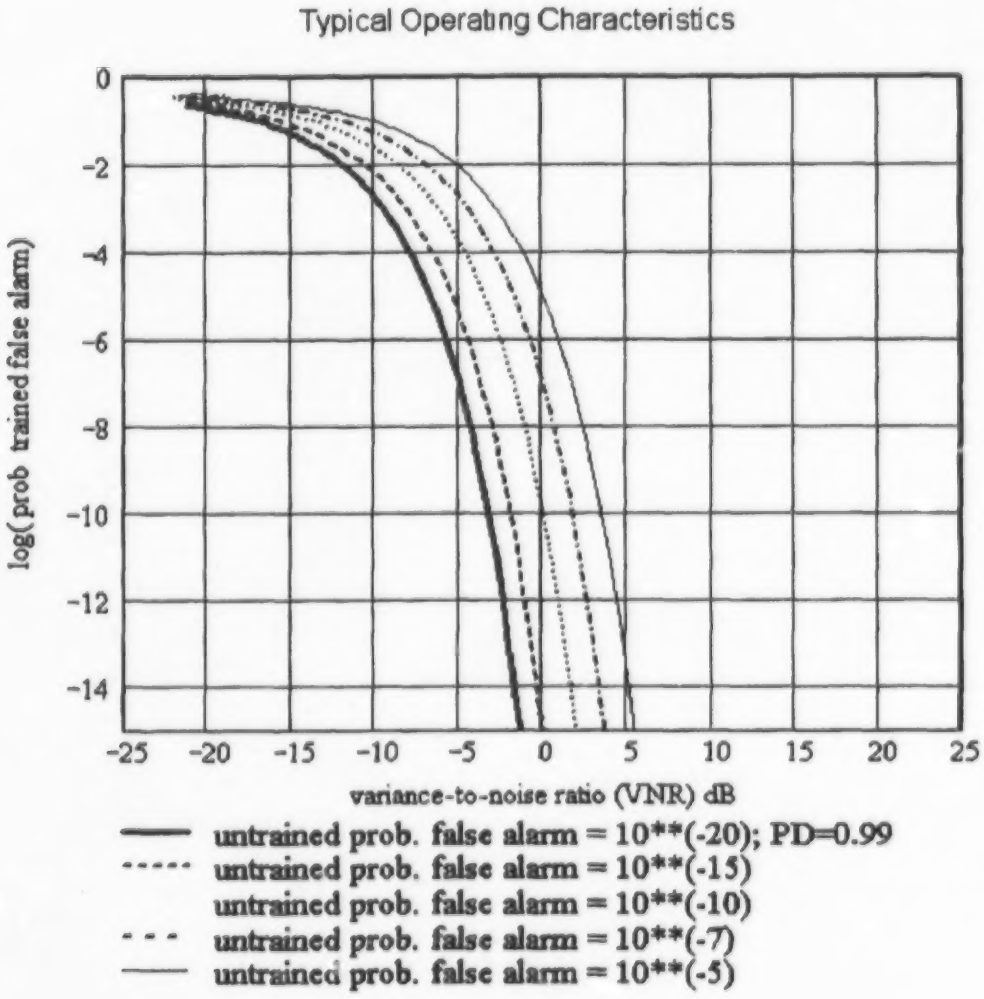

Fig. 4 Probability of trained false alarm
The probability of *trained* false alarm P_{TF} indicates how well rejection training will work. Typical values are plotted here as a function of the scanned image variance-to-noise ratio (*VNR*) in (54), assuming a probability of detection P_D=0.990, and probabilities of *untrained* false alarm as noted in the legend. The probability of failed rejection increases as the *VNR* becomes smaller

Fig. 5 The Esquimalt target range features a number of cylindrical "mine-like" pipes lying on their side, one standing vertically (bottom left), and one sphere (small, three-quarters down from top), in a relatively benign environment. The image is roughly 50 m wide by 260 m long. The multi-beam side scan sonar tow path runs from top to bottom, along the right side of the image, but off the image. Figure 6 is a magnified composite showing all targets clipped from this image.

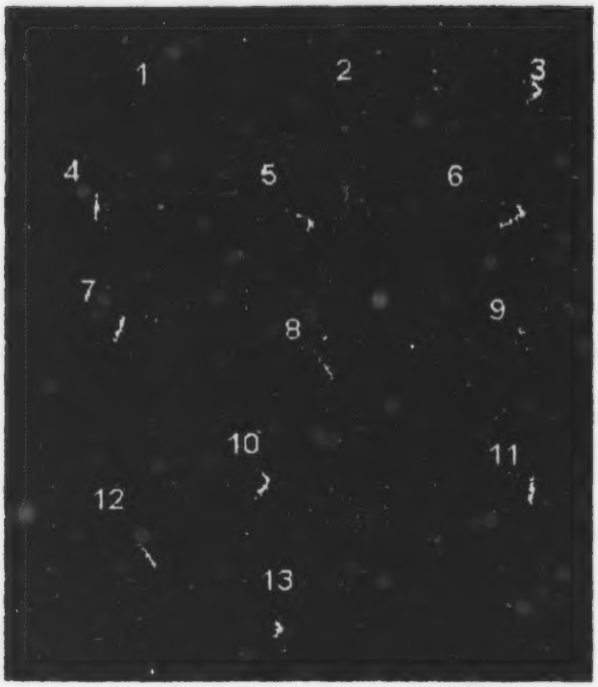

Fig. 6 Magnified composite of mine-like targets
Features whose highlight-shadow signatures are complete in the target range Fig.(5) have been clipped and pasted into a collage for better viewing for the examples to follow. They include:

1. unknown--large square scrap metal perhaps
2. resolution target 2.44X2.44m table with three 5 cm diameter spheres attached in 'V'-pattern
3. horizontal pipe, 0.8 m diameter, 2 m long
4. unknown
5. horizontal pipe, 0.8 m diameter, 2 m long
6. horizontal pipe, 0.8 m diameter, 2 m long
7. horizontal pipe, 0.8 m diameter, 2 m long
8. horizontal pipe, 0.7 m diameter. 2 m long
9. unknown--a rock perhaps
10. horizontal pipe, 0.8 m diameter, 2 m long
11. horizontal pipe, 0.8 m diameter, 2 m long
12. horizontal pipe, 0.7 m diameter. 2 m long
13. vertical pipe, 0.8 m diameter, 2 m long

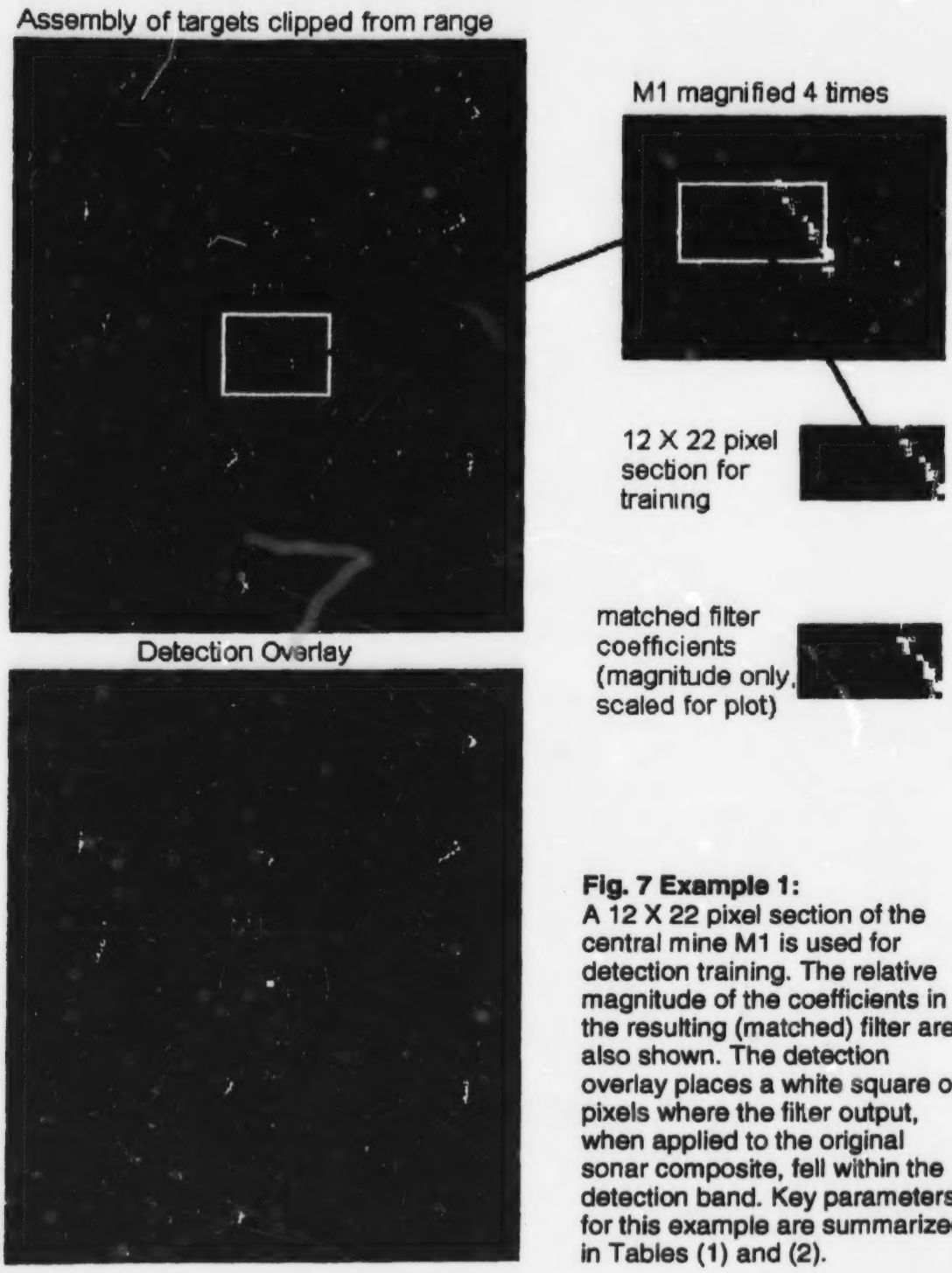
Assembly of targets clipped from range
M1 magnified 4 times
12 X 22 pixel section for training
matched filter coefficients (magnitude only, scaled for plot)
Detection Overlay

Fig. 7 Example 1:
A 12 X 22 pixel section of the central mine M1 is used for detection training. The relative magnitude of the coefficients in the resulting (matched) filter are also shown. The detection overlay places a white square on pixels where the filter output, when applied to the original sonar composite, fell within the detection band. Key parameters for this example are summarized in Tables (1) and (2).

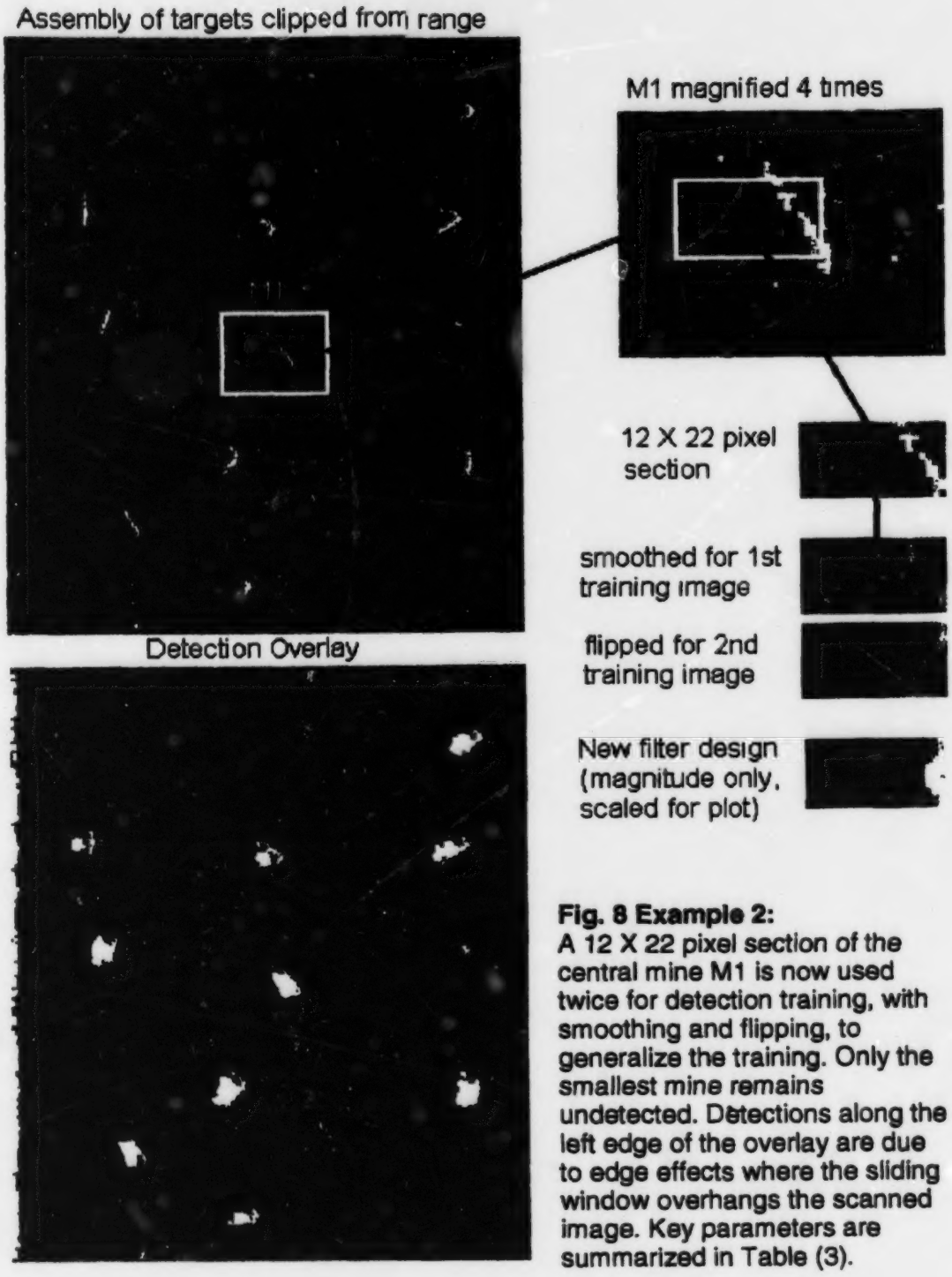
Assembly of targets clipped from range
M1 magnified 4 times
12 X 22 pixel section
smoothed for 1st training image
flipped for 2nd training image
New filter design (magnitude only, scaled for plot)
Detection Overlay

Fig. 8 Example 2:
A 12 X 22 pixel section of the central mine M1 is now used twice for detection training, with smoothing and flipping, to generalize the training. Only the smallest mine remains undetected. Detections along the left edge of the overlay are due to edge effects where the sliding window overhangs the scanned image. Key parameters are summarized in Table (3).

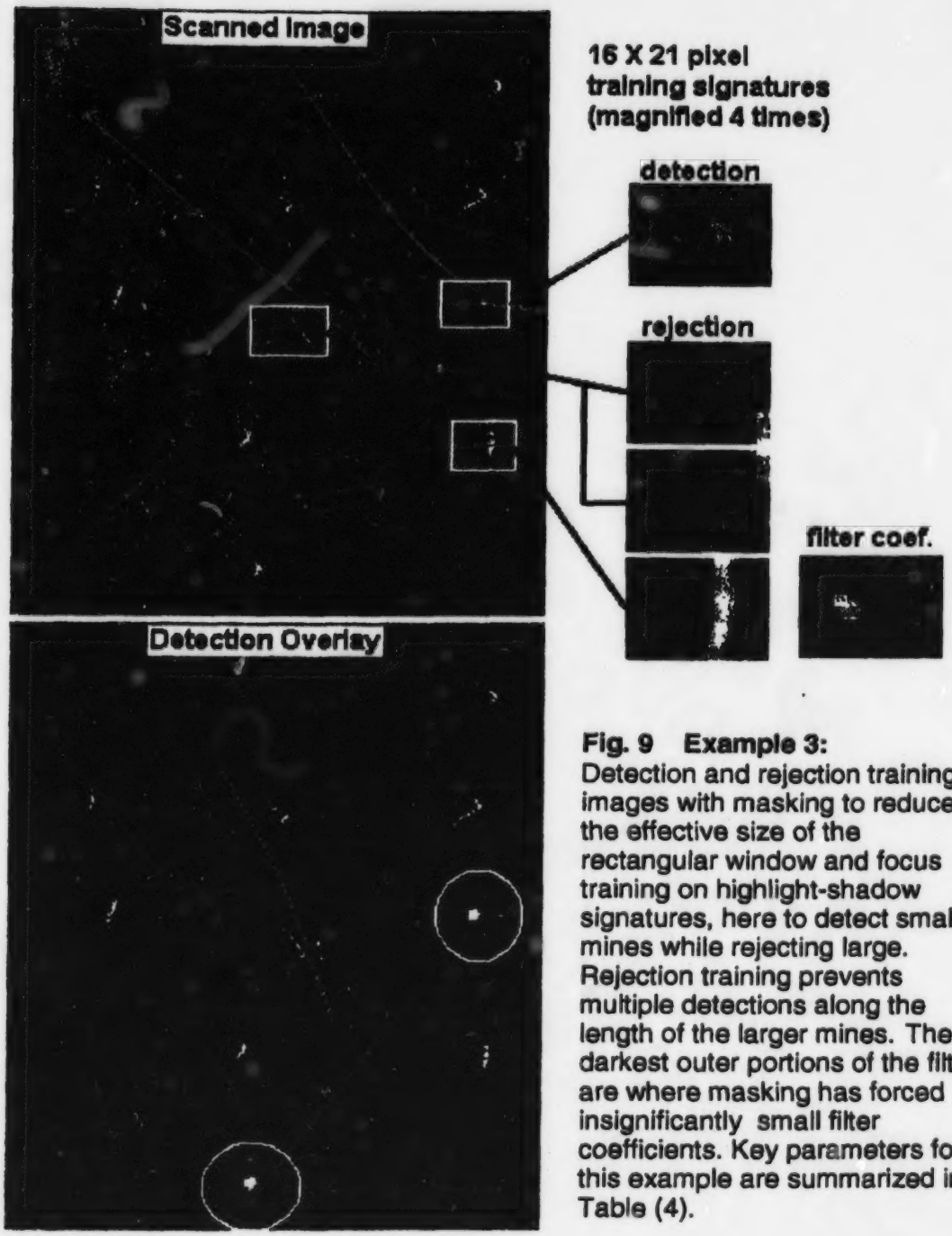

Fig. 9 Example 3:
Detection and rejection training images with masking to reduce the effective size of the rectangular window and focus training on highlight-shadow signatures, here to detect small mines while rejecting large. Rejection training prevents multiple detections along the length of the larger mines. The darkest outer portions of the filter are where masking has forced insignificantly small filter coefficients. Key parameters for this example are summarized in Table (4).

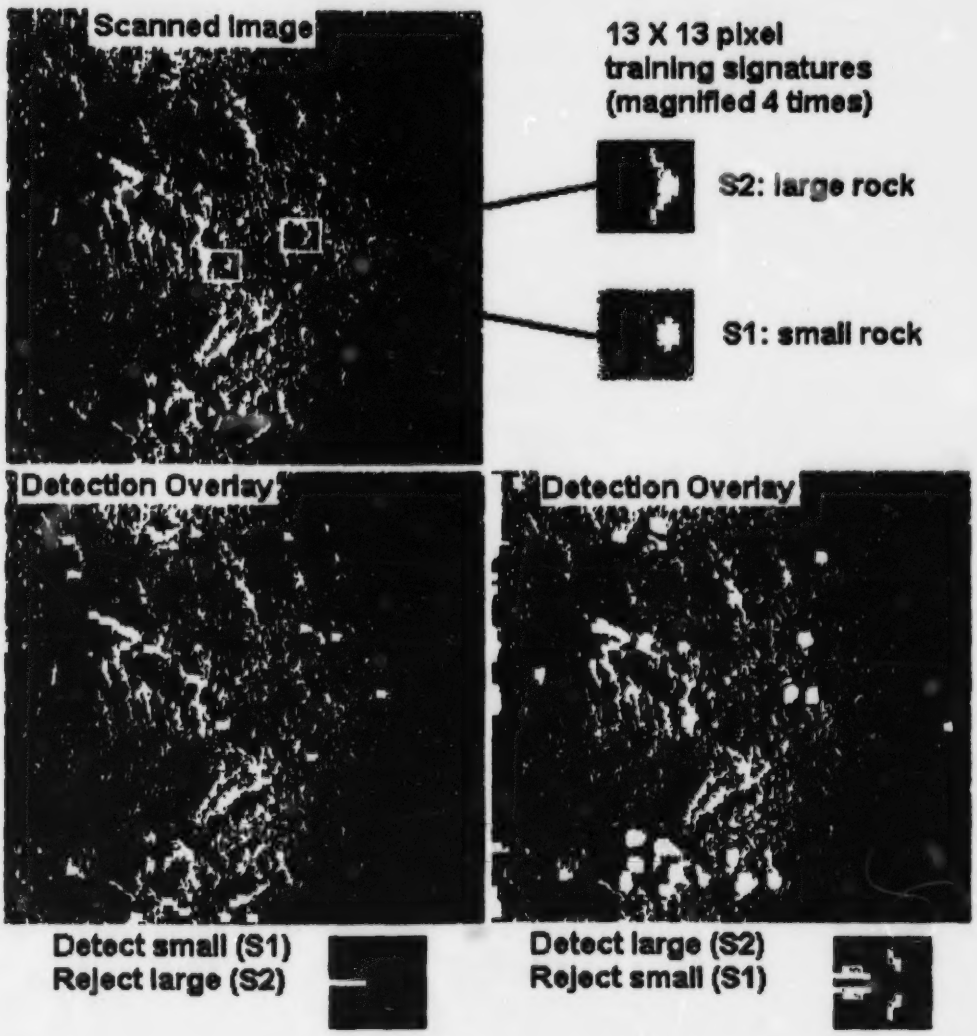

Fig. 10 Example 4: Discriminating between small and large rocks
Detection and rejection training used to distinguish rock size. The training images were based on the small and large rocks, S1 and S2, respectively. Masking was used to focus on the highlight-shadow signatures alone. The relative magnitude of the filter coefficients used for each rock size are shown below their respective detection overlays. The darkest outer portions of the filters are where masking has forced insignificantly small filter coefficients. Detections along the left edge for large rocks are due to edge effects, where the filter window overhangs the scanned image. Key parameters for this example are summarized in Tables (5) and (6).

UNCLASSIFIED
SECURITY CLASSIFICATION OF FORM
(highest classification of Title, Abstract, Keywords)

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1 ORIGINATOR (the name and address of the organization preparing the document Organizations for whom the document was prepared, e g Establishment sponsoring a contractor's report, or tasking agency, are entered in section 8)

Defence Research Establishment Atlantic
PO Box 1012
Dartmouth, Nova Scotia, Canada B2Y 3Z7

2 SECURITY CLASSIFICATION (overall security classification of the document including special warning terms if applicable)

UNCLASSIFIED

3 TITLE (the complete document title as indicated on the title page Its classification should be indicated by the appropriate abbreviation (S,C,R or U) in parentheses after the title)

Trainable linear filters for the automated detection of mines in side scan sonar images

4 AUTHORS (Last name, first name, middle initial If military, show rank, e g Doe, Maj John E)

Kessel, Ronald T

5 DATE OF PUBLICATION (month and year of publication of document)	6a NO OF PAGES (total containing information Include Annexes, Appendices, etc)	6b NO OF REFS (total cited in document)
Dec 1999	34 (approx.)	15

7 DESCRIPTIVE NOTES (the category of the document, e g technical report, technical note or memorandum If appropriate, enter the type of report, e g interim, progress, summary, annual or final Give the inclusive dates when a specific reporting period is covered)

TECHNICAL MEMORANDUM

8 SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development Include address)

Defence Research Establishment Atlantic, Electromagnetics Section, MCM Group

9a PROJECT OR GRANT NO (if appropriate, the applicable research and development project or grant number under which the document was written Please specify whether project or grant)	9b CONTRACT NO (if appropriate, the applicable number under which the document was written)

10a ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity This number must be unique to this document)	10b OTHER DOCUMENT NOs (Any other numbers which may be assigned this document either by the originator or by the sponsor)
DREA TM 1999-171	

11 DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification)

(X) Unlimited distribution
( ) Defence departments and defence contractors; further distribution only as approved
( ) Defence departments and Canadian defence contractors; further distribution only as approved
( ) Government departments and agencies, further distribution only as approved
( ) Defence departments; further distribution only as approved
( ) Other (please specify).

12 DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document This will normally correspond to the Document Availability (11) However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected)

UNCLASSIFIED
SECURITY CLASSIFICATION OF FORM
UNCLASSIFIED
SECURITY CLASSIFICATION OF FORM
(highest classification of Title, Abstract, Keywords)

13 ABSTRACT (a brief and factual summary of the document It may also appear elsewhere in the body of the document itself It is highly desirable that the abstract of classified documents be unclassified Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U) It is not necessary to include here abstracts in both official languages unless the text is bilingual)

(U) The multibeam side scan sonar images used for sea mine hunting look rather like aerial photographs of the seafloor, but there are important differences. The sonar's automatic gain control, the consistent highlight-shadow signature progression of objects, and the absence of vanishing points, considerably reduce the degrees of freedom in an object's appearance, making sonar images better suited for linear detection schemes than optical photographs generally are This paper demonstrates a trainable linear filter for the detection of mine-like objects in side scan sonar images The training is fast and flexible, using actual or synthetic sonar images to prescribe what image patterns are to be detected and rejected, and with a sliding filter window of any shape to focus more particularly on important features Measures of performance such as the probabilities of detection and false alarm, and the limits of trained discrimination and window size are derived. The method is demonstrated for mine-like targets in actual sonar images.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document They should be selected so that no security classification is required Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included If possible keywords should be selected from a published thesaurus e g Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified If it not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title)

mine hunting, side scan sonar, image processing, detection